Characterizing QoE in Large-Scale Live Streaming

Thiago Guarnieri
Ítalo Cunha
Jussara Almeida
Universidade Federal de Minas Gerais
{*thiago, cunha, jussara*}@*dcc.ufmg.br*

Idilio Drago
Politecnico di Torino
idilio.drago@polito.it

Alex B. Vieira
Univ. Federal de Juiz de Fora
alex.borges@ufjf.edu.br

***Abstract*—Understanding the impact of performance degradation on users' QoE during live Internet streaming is key to maximize the audience and increase content providers' revenues. It is known that some problems have a strong correlation with low QoE—e.g., users experiencing video stalls tend to leave video sessions earlier. It is however mostly unknown whether such observations hold for live streaming of large-scale events (e.g., the FIFA World Cup). Such events are particular due to the attractiveness of the streamed content, reaching an impressively high audience worldwide. We study whether and to what extent performance degradation during live streaming of large-scale events affects users' QoE. We leverage a unique dataset collected from a major content provider in South America during the 2014 World Cup. We first extract performance metrics from the logs: stream bitrate and bitrate switches, playback stalls, and playback startup latency. We then correlate these performance metrics with session duration, which we use as a QoE indicator. We confirm the strong correlations between the metrics and QoE indicators; in particular, frequent stalls are often accompanied by higher probability of early session termination. Moreover, we quantify how such correlations vary according to broadcasted matches and client terminals. Some of our findings challenge intuition – e.g., we find that PC users seem more tolerant to problems than users on mobile terminals. Our results and dataset are an important step towards models to predict users' QoE in large-scale events.**

I. Introduction

Despite streaming traffic growth and commercial success of on-demand streaming services, *live* streaming of large-scale events is still frequently accompanied with performance and scalability problems [1], [2], which lead to poor user Quality of Experience (QoE). Understanding causes of QoE degradation is important to content providers, who have to provision costly large-scale distribution infrastructures, and to end-users, who need to compare and choose service providers.

Studying live streaming QoE is challenging because it is subjective: QoE depends on the user, his expectations, context, and on the content being transmitted. Another challenge is that data that captures QoE is seldom available. Finally, ground truth requires surveying users about their experience, which is costly and does not scale. Although there has been recent progresses in better understanding user QoE in live streaming, the available knowledge is still limited [3], [4], [5].

In this paper, we provide a study of the live streaming of 2014 FIFA World Cup by a major content provider in South America (§II). Our analyses rely on logs from more than 20 million user sessions collected from HTTP Web servers (§III). Although HTTP request logs do not contain QoE information, we model and emulate the streaming client's local state to compute QoE-related metrics [3], [6]: average session bitrate, playback startup latency, playback stalls, and bitrate adaptation events (§IV). Our method can be applied to other datasets of HTTP request logs to further expand our understanding about user QoE for different content and user bases.

We then correlate the aforementioned performance metrics with session duration, which we use as a proxy for user engagement and QoE (§V). Our results cover client sessions with diverse performance metrics. We observe that a significant fraction of sessions experience a high rate of stalls and bitrate adaptations, e.g., about 20% of sessions suffer at least one playback stall per minute. We quantify the impact of performance metrics and find that users are more willing to tolerate long playback startup delays and low bitrates than playback stalls or frequent bitrate degradation. This observation generalizes across different soccer matches as well as across PC and smartphone clients. Finally, we find that PC users are more tolerant to poor performance than smartphone users and watch the stream for longer periods under the same performance conditions.

II. Streaming system and dataset

HTTP live streaming infrastructure. We analyze HTTP live streaming server logs collected by Globo.com, a major content provider in South America. Globo.com delivers video using Apple HTTP Live Streaming (HLS) [7]. Globo.com employs a standard video distribution architecture, illustrated in Fig. 1. Media coming from the recording devices are encoded in different bitrates, split into small segments, and indexed. Index files (.m3u8) aggregate a list of segments and contain metadata about the available bitrates and paths to segments. All index and media files are then transmitted to two content distribution centers. Each distribution center runs Web servers to receive and process standard HTTP requests. Clients first request segment index files and, then, keep downloading segments and further index files to receive the media. Clients continuously estimate network performance to decide at which bitrate to request the next segment.

Globo.com distributes media from two distribution centers, one in São Paulo and another one in Rio de Janeiro, the first and second largest cities in Brazil, respectively. Globo.com peers with several ISPs at each center, and both centers are

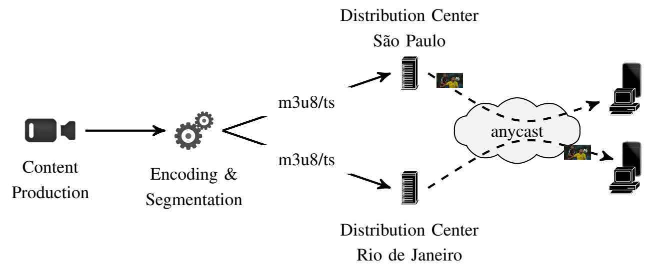


Fig. 1: Globo.com's live streaming architecture.

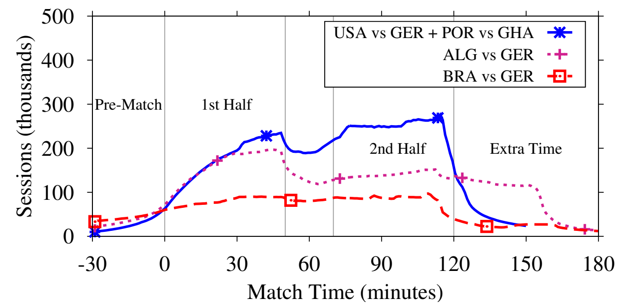


Fig. 2: Simultaneous sessions during four matches.

directly connected to Internet exchange points. Globo.com announces the same set of IP prefixes from both centers and clients watching a stream reach one of the distribution centers transparently via anycast.

Dataset. We analyze the set of HTTP request logs collected at Globo.com's live streaming servers in both distribution centers during the 64 matches of the 2014 FIFA World Cup. Globo.com was the official TV and online broadcaster of the event in Brazil, so we expect the request logs to capture the majority of Brazil's online audience of the soccer World Cup.

Every HTTP request for a video fragment is logged by Globo.com's nginx servers in the default logging format, containing request date, time, the client IP address, requested URL, HTTP status, bytes sent, and the user agent string. Segment URLs contain the requested channel (i.e., the match) and the segment encoding bitrate. Most proxies add an X-Forwarded-For HTTP header in the request, which we leverage to differentiate users. Finally, browser user agent strings allow us to identify device types (e.g., PC, tablet, or smartphone).

Infrastructure and client configuration. Globo.com encodes media in up to six bitrates varying between 264–2564 kbps.[1] We identify bitrate of requested segments from URLs. The encoder generates segments (.ts files) that contain 3 seconds of playback.

We identify clients using their public IP address and user agent string; we also use the private IP addresses of clients behind proxies whenever this information is available. We observe that clients request 98.9% of consecutive segment indexes and media segments within intervals shorter or equal to than 3 seconds. Given this behavior, we conservatively define a streaming *session* as a sequence of requests from the same client that are not separated by an interval longer than 120 seconds. We have evaluated different threshold values (from 30 s to 180 s) yielding qualitatively similar results.

NATs could interfere with our analysis: when two or more clients behind a NAT, using exactly the same browser version (i.e., user agent string), watch a single match, we would mark all their requests as belonging to the same session. NATs are however easily recognizable in our data. Indeed, since each client behind the NAT must perform a request to retrieve every video fragment, NATs result in sessions with clearly abnormal numbers of duplicated requests for fragments. Less than 5% of sessions in our dataset present such behavior.

[1] In some occasions with more than one match happening simultaneously the highest bitrate was restricted to users of particular devices (e.g., mobile phones) likely to limit server workloads.

III. Workload overview

Overall, we observe up to 1.1 million unique clients in a single day with 4 matches, and up to 470 thousand simultaneous sessions during a single match (Argentina vs Switzerland). We find no clear correlations between the *appeal* of a match (e.g., a semi-final/final match or a match involving Brazilian national team) and the total number of user sessions. This may be due to all matches also being broadcasted on open television as well as happening at different times and days of the week. For example, users may prefer to watch more appealing games on TV, which would impact the online audience.

By parsing client user agent strings, we identify that 81% of sessions are started from PCs, 12% from smartphones, and 6% from tablets. A few sessions from unidentified device types and robots complete the figure. We find a significant fraction of sessions from mobile devices, which have not been discussed explicitly in previous work and might have different trade-offs for user QoE. In Sections V and VI, we focus our characterization on sessions from PCs and smartphones.

Fig. 2 shows the number of simultaneous user sessions during four representative match transmissions that we use as case studies throughout the paper: USA vs Germany, Portugal vs Ghana, Algeria vs Germany, and Brazil vs Germany. The figure shows the number of simultaneous sessions over time, staring 30 minutes before the beginning of the matches until 1 hour after normal time is over. The blue solid line shows the number of sessions for two simultaneous games. There is a decrease in the number of sessions during the 15-minute half-time, and matches with extra time (e.g., ALG vs GER) show high audience for the additional 30 minutes. We choose these high-interest matches as representative because they capture behavior observed in other matches, including a match with extra time and a transmission of simultaneous matches.

Fig. 3 shows the empirical Cumulative Distribution Function (CDF) of the duration of streaming sessions during the World Cup. Fig. 3(a) characterizes the duration of sessions over our selected four matches, whereas Figs. 3(b) and 3(c) show session duration per device type for different matches.

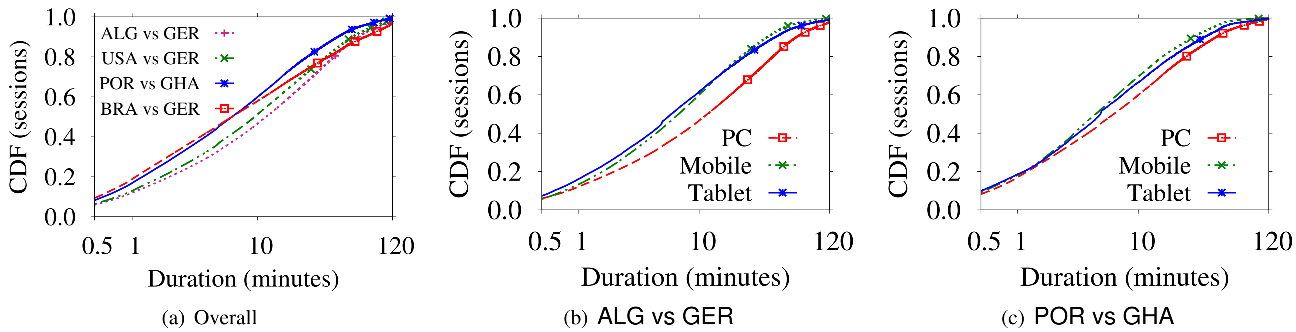


Fig. 3: Session duration per match and device type.

Overall, sessions are relatively short, with median session time between 8 and 12 minutes, depending on the match. Matches with extra time (e.g., ALG vs GER) have longer sessions, as expected. Interestingly, around 20% of the sessions are abandoned in less than 1 minute. We will show later that shorter sessions have longer startup delays, hinting at the (negative) correlation between startup delay and user engagement.

Figs. 3(b) and 3(c) show that mobile phones and tablets have shorter sessions than PCs. Fig. 3(c) shows the duration of sessions for POR vs GHA, broadcasted simultaneously with USA vs GER. A smaller difference is noticed among device types compared to other games (e.g., ALG vs GER in Fig. 3(b)). Manual inspection shows a fraction of PC users alternate between the simultaneous matches, thus reducing overall PC session duration in such cases.

IV. Correlating performance metrics and user Quality of Experience

We compute the following QoE-related metrics [3], [6]:

- *Average bitrate.* Image quality is fundamental to QoE in live streaming and can be estimated by the average bitrate. We extract each requested segment's bitrate from the URLs, and compute each session's average bitrate over its duration.
- *Playback startup delay.* The time elapsed from a user accessing the live stream and the actual beginning of playback. Measuring startup delays from the server side is challenging as we lack client-side status information. We estimate startup delay for a client measuring the time from the first request for an index file to the arrival of a segment request after enough segments to fill the buffer have been served. This overestimates the startup delay when clients begin playback before the buffer fills or when clients do not request the next media segment immediately after starting playback. In the latter case, startup delay overestimation error is on the order of one segment playback time, and the relative error is inversely proportional to buffer size.
- *Playback stalls.* We estimate the rate of stall events as the number of buffer underflows divided by session duration. We use the segment playback time and buffer size to emulate the client and estimate its buffer occupation over time. We assume clients receive segments immediately after requests are satisfied (i.e., appear in server logs). This assumption is conservative and overestimates the client's buffer level and underestimates playback stalls. In practice, we expect the impact to be low as chunk playback time (3 seconds) is, at least, one order of magnitude longer than usual round-trip times.
- *Rate of bitrate adaptation events.* We evaluate how the bitrate evolves during a session. Even though reducing the bitrate helps prevent playback stalls, their occurrence can still negatively impact QoE as they impact the quality of the video. We identify bitrate reduction events whenever two consecutive segment requests in the streaming session have decreasing bitrates. Note that bitrate reductions often occur after playback stalls, while the buffer recovers. We do not consider such reductions as bitrate adaptation events to avoid counting them twice; i.e., we consider them as playback stall events. When a client requests a segment at multiple bitrates, we assume the lower bitrate is played back if it prevents a playback stall compared to waiting for the higher bitrate chunk.

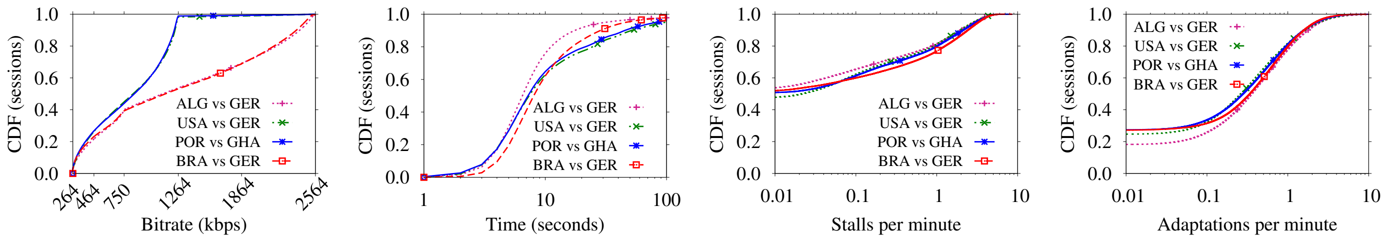
We first characterize the four metrics independently, giving insights into the overall workload on the servers and user QoS. We then correlate each performance metric with normalized session duration, which we use as a proxy for user engagement. We consider Spearman correlation as the data we analyze have nonlinear relationships. Spearman correlation is a non-parametric test based on ranks, which indicates how well the relation between two variables can be described by a monotonic function.

V. Metrics and engagement on PCs

We first investigate performance metrics and their impact on client session times on PC clients. We compare results to mobile clients in Section VI.

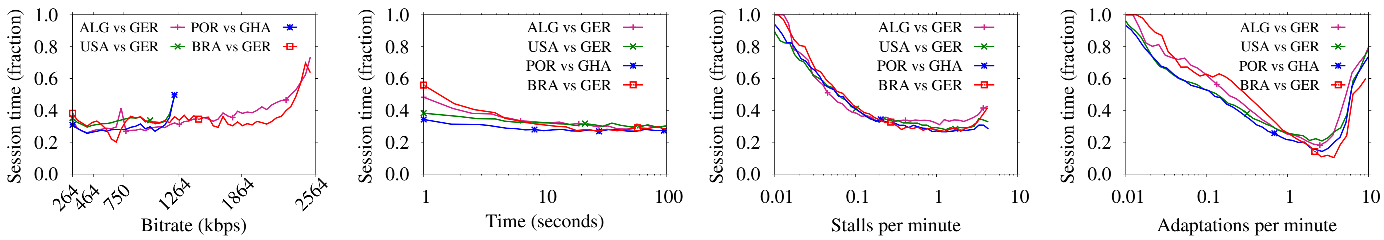
A. Performance evaluation

Figure 4 shows the CDFs of our performance metrics for the selected four matches. Figure 4(a) shows how client sessions experience a wide range of average transmission bitrate. Lines for USA vs GER and POR vs GHA, as well



(a) Mean Bitrate (b) Start-up Delays (c) Rate of Stalls (d) Rate of Adaptations

Fig. 4: Performance metrics for PC sessions.



(a) Mean Bitrate (b) Start-up Delays (c) Rate of Stalls (d) Rate of Adaptations

Fig. 5: Impact of performance on normalized session duration (PC users).

as for ALG vs GER and BRA vs GER overlap. The difference between the two curves is caused by changes to the provider's setup. For example, the maximum available bitrate was reduced to 1264kbps for USA vs GER and POR vs GHA, streamed simultaneously. Possible reasons include capacity management at the provider's infrastructure. All matches show a large variation on the average bitrate among different sessions, likely due to varying network conditions and last-mile bandwidth bottlenecks. For matches that have the same available bitrates, average bitrates are similar across different matches. Figure 4(a) also shows that average bitrate is not highly concentrated around the encoding bitrates used by the provider (marked in the x-axis). This behavior is explained by a large number of video bitrate adaptations shown in Figure 4(d), discussed below.

Figure 4(b) shows that a significant fraction of clients have large session playback startup delays. Around 30%–40% of the sessions have startup delays higher than 10 seconds. Figure 4(c) shows that clients often experience high playback stall rates: at least half the sessions have more than one stall, and around 20% of sessions have at least one stall per minute of playback. Finally, Figure 4(d) shows the rate of bitrate adaptations. Around 70%–80% of the sessions have no more than one bitrate adaptation event per minute. Generally, sessions experience more bitrate adaptation events than stalls (compare distributions in Figures 4(c) and 4(d)). This suggests that adaptations prevent a significant number of stalls.

Even though the analyzed matches occur at different dates, times, and impose different workloads on the provider (see Figure 2), our observations on streaming performance generalize across all four matches as well as across the other 60 matches in our dataset (omitted for brevity).

B. Impact of performance on session duration

We analyze how the four performance metrics above correlate with session duration. As in previous work [3], we use session duration as a proxy of user engagement and QoE. Figure 5 presents the impact of the performance metrics on session duration for PC clients. In the graphs, we have discretized the x-axis in 50 bins. For mean session bitrate (Figure 5(a)), we generate bins of equal size. For the other metrics (Figures 5(b)–5(d)), we applied logarithmic binning – i.e., exponentially increasing bin sizes so that each bin has the same width on the logarithmic x-axis. We normalize session duration by the remaining match time, i.e., a value of 1 means that the user watched the match to its conclusion. We compute the mean normalized session duration for the sessions in each bin. Note that the number of sessions in each bin varies (and is given in the CDFs in Figure 4). To avoid bias from short sessions starting at the end of the match, which would have high normalized session duration, we consider only sessions starting up to 15 minutes before the end of matches. Again, we show four matches results as a matter of illustration, with most of the remaining matches leading to similar plots.

Results show a clear non-linear correlation among performance metrics and session duration. Some metrics, however, seem to influence user engagement more decisively than others do. For example, Figure 5(a) shows that the fraction of the match that is watched in a session increases when users experience mean bitrate close to the maximum rate offered by the provider in the match. Note the spikes at around 1264 kpbs for POR vs GHA and around 2564 kpbs for ALG vs GER.

Sessions with the best possible quality stay active in general up to 70% of the remaining match time, whereas sessions experiencing lower quality stay active for less than 40% of the remaining match time.

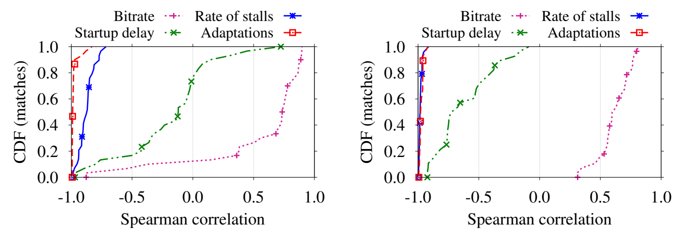
Similar impact can be seen for other metrics, at different intensities. Figure 5(b) presents the relation between mean startup delays and session duration. Clients on sessions with long startup delays will leave the streaming (on average) earlier than sessions with near-instantaneous startup. The decrease on engagement is however less intense than what is observed for the mean bitrate. Sessions that start after 10 seconds typically will stay active for 30% of the remaining match time, whereas sessions with 1 second startup delay will be active for more than 50% of the remaining match time.

More dramatic (negative) correlations can be observed for the remaining metrics. Focusing on Figure 5(c), notice how the fraction of the remaining match time that is watched in a session sharply decreases when the rate of playback stall increases. The figure stops at five stalls per minute, since sessions with higher stall rates are extremely rare. In fact, since each fragment carries 3 seconds of video in this deployment, larger stall rates would imply that the playback is stuck during most of the session. More interestingly, sessions with 1 stall per minute stay on-line only 40% of the remaining match time, whereas sessions with negligible number of stalls watch most of the match after joining the streaming. Notice that, similarly to [3], we observe a significant mass of sessions with very good quality and low duration (see Figure 4(c)). Users joining the streaming only briefly may have low interest in the matches, and bias session duration for the group of users with zero stalls. We omit these cases from Figure 5(c).

The rate of streaming bitrate adaptations also strongly correlates with session duration (Figure 5(d)). We see a sharp decrease in the fraction of match time that sessions remain active when contrasting sessions with 1 adaptation per minute against sessions with a negligible bitrate adaptation rate. The close relation between stalls and bitrate adaptation can explain this result. In fact, users with a high number of adaptations are likely under poor network conditions, which may eventually lead to stalls.

Notice that we observe an increase in the fraction of match time covered by the sessions that are marked with 5–10 adaptations per minute. In fact, this is a measurement artifact in our data. Note that each fragment carries 3 seconds of video and thus 10 adaptations per minute is the maximum possible value in Globo.com's setup. Those sessions (around 3% of the total) are the cases where two or more users, relying on exactly the same browser and behind a NAT, watch the same match at different bitrates. Since our methodology mixes the traffic of these users, we erroneously identify adaptations. We ignore these few data points in the following results.

Spearman correlation coefficients corroborate the visual intuition from the figures. Figure 6(a) summarizes the correlation coefficients calculated for 32 different matches for PC clients. CDFs of the correlation coefficients for different metrics are presented. Results show that startup delays are



(a) PC

(b) Mobile

Fig. 6: Spearman correlation coefficients between metrics and session duration for PC and mobile users.

the least correlated to user engagement, i.e., Spearman coefficients are usually closer to zero than in other metrics. This result suggests that users are willing to wait for the video to start playing when watching it on PCs.

Next comes the streaming bitrate, even if we observe divergent results in some few matches—e.g., very negative correlations. Such variation happens because the content provider artificially set the maximum bitrate to low values in some matches, to handle larger workloads, thus biasing results.

Finally, stalls and bitrate adaptations have the most significant impact on user engagement, indicating users are less willing to tolerate stalls and bitrate degradation.

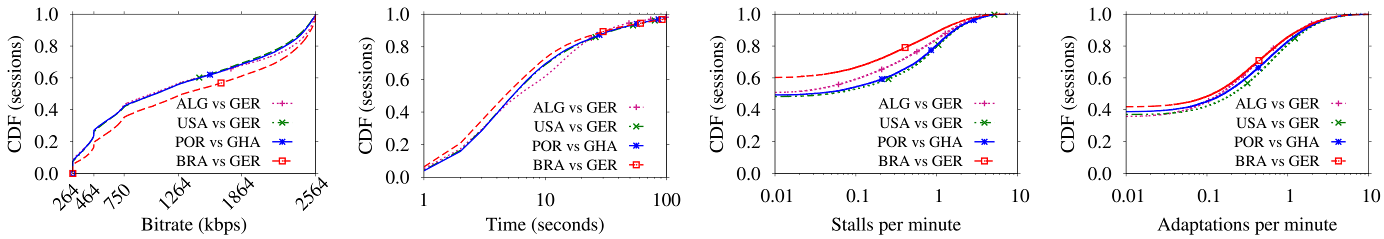
VI. Metrics and Engagement on Smartphones

Figure 7 shows the distributions of performance metrics focusing now only on sessions from smartphones. The figure includes precisely the same metrics as in the previous section, calculated following the same methodology. Number and limits of the bins are maintained as well.

We highlight that CDFs for metrics in mobile clients are generally similar to those observed for PCs. Major differences are seen in the case of mean bitrate (Figure 7(a)). We observe that the provider has not applied bitrate restrictions to mobile clients during the selected matches. Moreover, we can observe slightly lower percentages of users with stalls and bitrate adaptations on mobile phones than on PCs–compare Figures 7(c) and 7(d) with their equivalents in Figure 4.

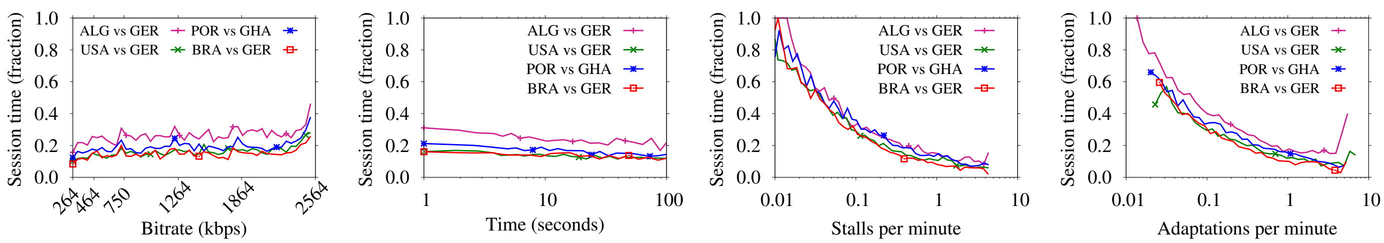
More interesting, mobile devices present much stronger correlations and lower variance among different matches between performance metrics and normalized session duration, when compared to PC clients. Compare curves in Figures 8(a)–8(d) to those in Figures 5(a)–5(d). Note also that artifacts seen for the rate of adaptations on PCs are almost not present in Figure 8(d), as the number of smartphones is much lower than the number of PCs in our dataset—i.e., the odds of finding multiple smartphones in NATs are reduced.

The intensity of variations in normalized session duration follow similar pattern in PCs and mobile phones when we consider mean bitrate and start-up delays. That is, as bitrate decreases, or as the start-up delay increases, we observe a respective decrease in mean normalized session duration. We will see next that correlations for these metrics are stronger



(a) Mean Bitrate (b) Start-up Delays (c) Rate of Stalls (d) Rate of Adaptations

Fig. 7: Performance metrics for mobile phone sessions.



(a) Mean Bitrate (b) Start-up Delays (c) Rate of Stalls (d) Rate of Adaptations

Fig. 8: Impact of performance on normalized session duration (mobile users).

with mobile devices than with PCs, thanks to the more homogeneous numbers across matches.

Users in mobile devices seem much more impacted by stalls and bitrate adaptations than those watching from PCs. Observe that mobile sessions with one stall per minute on average will stay active less than 20% of the remaining match time. Compare these figures with numbers for mobiles sessions with a negligible rate of stalls. Compare them also to numbers for sessions on PCs. Sessions with one stall per minute on PCs stay active twice as much as sessions on mobile phones with similar rate of stalls.

Correlation coefficients for this scenario are shown in Figure 6(b) and, confirm the visual intuition. Comparing results of mobile phones to those obtained for PCs, we confirm that users in mobile phones are even more sensitive to performance variations. Indeed, we not only can visually notice that the correlations are stronger in mobile phones than in PCs, but also we can see that Spearman correlation coefficients are larger in term of absolute values for all metrics.

This last result may sound counter-intuitive, since users should be used to lower quality of connections in mobile terminals when compared to PCs. Thus, one could expect mobile users to be more tolerant to problems, in contrast to what our data show. We conjecture that mobile users are less tolerant to problems because of the way the content is consumed. We expect that a significant portion of PC users leave the match playing in background – e.g., listing to the transmission while performing other tasks. Mobile users instead actively engage on watching the match, thus reacting to problems faster. We plan to explore these aspects further in future work.

VII. Related Work

The work of Dobrian et al. [3] is the closest to ours. The authors present a large-scale QoE study based on client-side instrumentation. They evaluate over 2 million views from more than 1 million users, correlating quantitative metrics to user engagement. They find that the rate of stalls has the largest correlation with user engagement across all analyzed video contents. Moreover, the average bitrate at which content is streamed correlates better with playback time on live video than on video on-demand.

Our work extends Dobrian et al.'s analysis. We provide a thorough evaluation of QoE-related metrics in large-scale streaming events, which the previous work only preliminarily considered. We confirm that many of Dobrian et al.'s findings hold for large-scale events as well. For example, we also find that the rate of stalls has a strong correlation with session duration. Beyond their findings, we show correlations between bitrate switches and session duration, and explore how the several metrics vary across the type of user device, e.g., concluding that mobile users seem to be more sensitive to problems than users watching videos on PCs.

Krishnan et al. [8] also perform a deep analysis of QoE-related metrics on video streaming. They try to establish a causal relationship between video quality and viewer behavior, thus taking a step beyond purely correlational studies. For example, they not only confirm the correlation between rate of stalls and abandonment rate, but also show evidences of causality. These authors however do not address large-scale live streaming as we do. Applying their methodology to search for causalities in our dataset is a natural continuation for our work, which we will pursue in future work.

Other works considering large-scale live streaming events are scarce [5], [6]. Yin et al. [9] analyzed the 2008 Olympic Games, a large-scale event that attracted on-line audiences unseen up to that period. Similarly, authors of [10] evaluate the 2013 Super Bow. Both works focus on users' access patterns, particularly on the effects of flash crowds and user behavior, without considering QoE-related metrics.

We are aware of other works addressing QoE of users on video sessions [11], [12], [13]. In most cases, authors either perform analyses of small datasets collected in the network based on passive methodologies, or focus on ordinary live and on-demand video providers [14]. Our work is orthogonal to those efforts, since we focus on QoE-related metrics on large-scale events, which present very peculiar characteristics.

QoE is critical in Internet video applications as it impacts revenues for content providers and delivery systems. Our evaluation has implications for content providers, which can better plan resources for maximizing user engagement. In this direction, Yin et al. [15] propose a novel predictive control algorithm that optimally combines throughput and buffer occupancy to select the bitrate that maximizes users' QoE. Our results can foment such analytic researchers and help to develop new adaptation algorithms aiming at improving overall QoE on large-scale streaming.

VIII. Discussion and Conclusions

Large-scale live Internet streaming attracts millions of viewers spread all over the world. Despite the importance of QoE to operators and clients, our understanding of user QoE is still limited. This paper improves our understanding of the impact of performance metrics on user QoE using a methodology that requires only HTTP server logs. Our study relies on logs collected from Globo.com, a major content provider in South America, during the 2014 FIFA World Cup.

Our results showed how several performance metrics correlate with an indicator of user QoE (i.e., session duration). In particular, we found that a significant fraction of sessions experience long startup delays, frequent playback stalls, and bitrate degradation in the analyzed content provider. We then quantified the correlation of such performance metrics with session duration.

We found that session startup delay has the lowest (absolute) correlation with session duration, indicating users are willing to wait for playback to start. Better streaming bitrate has positive correlation with session duration. However, we found evidences that frequent playback stalls and bitrate degradation events, typical in congested or under-provisioned networks, have a strong impact on session duration.

Some of our results are counter-intuitive at a first glance. For example, our data suggest that users are more tolerant to performance degradation when watching live videos of popular events in PCs. Investigating root-causes of such differences is planned for our future work.

References

[1] "CNN: Inauguration P2P Stream a Success, Despite Backlash," 2009, https://gigaom.com/2009/02/07/cnn-inauguration-p2p-stream-a-success-despite-backlash/.

[2] "ESPN's Live Streaming Is Buckling Under The Pressure Of The US World Cup Game," 2014, http://www.businessinsider.com/espn-down-during-world-cup-2014-6.

[3] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, and H. Zhang, "Understanding the Impact of Video Quality on User Engagement," in *Proceedings of the SIGCOMM*, 2011, pp. 362–373.

[4] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "Developing a Predictive Model of Quality of Experience for Internet Video," in *Proceedings of the SIGCOMM*, 2013, pp. 339–350.

[5] P. Juluri, V. Tamarapalli, and D. Medhi, "Measurement of Quality of Experience of Video-on-Demand Services: A Survey," *Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 401–418, 2016.

[6] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hobfeld, and P. Tran-Gia, "A Survey on Quality of Experience of HTTP Adaptive Streaming," *Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 469–492, 2015.

[7] "HTTP Live Streaming (HLS) - Apple Developer, 2016," http://developer.apple.com/streaming.

[8] S. S. Krishnan and R. K. Sitaraman, "Video Stream Quality Impacts Viewer Behavior: Inferring Causality Using Quasi-experimental Designs," in *Proceedings of the IMC*, 2013, pp. 211–224.

[9] H. Yin, X. Liu, F. Qiu, N. Xia, C. Lin, H. Zhang, V. Sekar, and G. Min, "Inside the Bird's Nest: Measurements of Large-scale Live VoD from the 2008 Olympics," in *Proceedings of the IMC*, 2009, pp. 442–455.

[10] J. Erman and K. K. Ramakrishnan, "Understanding the Super-Sized Traffic of the Super Bowl," in *Proceedings of the IMC*, 2013, pp. 353–360.

[11] P. Casas, M. Seufert, and R. Schatz, "YOUQMON: A System for On-line Monitoring of YouTube QoE in Operational 3G Networks," *SIGMETRICS Perform. Eval. Rev.*, vol. 41, no. 2, pp. 44–46, 2013.

[12] R. Huysegems, B. D. Vleeschauwer, K. D. Schepper, C. Hawinkel, T. Wu, K. Laevens, and W. V. Leekwijck, "Session Reconstruction for HTTP Adaptive Streaming: Laying the Foundation for Network-based QoE Monitoring," in *Proceedings of the IWQoS*, 2012, pp. 15:1–15:9.

[13] R. Schatz, T. Hossfeld, and P. Casas, "Passive YouTube QoE Monitoring for ISPs," in *Proceedings of the IMIS*, 2012, pp. 358–364.

[14] M. Trevisan, I. Drago, and M. Mellia, "Impact of Access Speed on Adaptive Video Streaming Quality: A Passive Perspective," in *Proceedings of the Internet-QoE*, 2016, pp. 7–12.

[15] X. Yin, A. Jindal, V. Sekar, and B. Sinopoli, "A Control-Theoretic Approach for Dynamic Adaptive Video Streaming over HTTP," in *Proceedings of the SIGCOMM*, 2015, pp. 325–338.